

살아있는 중국어 교재를 만드는 LLM

— 소셜미디어 텍스트의 교육적 변용 —

박민준*

<目次>

- | | |
|--------------------------------------|-------------------|
| I. 머리말 | IV. 논의 및 활용 사례 탐구 |
| II. 소셜미디어 텍스트의 특성과 교육 활용도 | V. 맺음말 |
| III. ChineseRealTalk 인터랙티브 교육 플랫폼 구현 | |

<중문초록>

本研究探討利用大語言模型(LLM)將中文社交媒體文本轉化為教學資源的方法，提出了一項交互式教學平台 *ChineseRealTalk*。在中文教學中，教材語言與母語者在社交媒體平台上使用的真實語言之間存在一定差距。儘管社交媒體文本能夠反映當代學生的語言表達習慣並具有較高的教學價值，但將其引入課堂仍面臨現實困難：教師需要根據學習者的中文水平手動調整難度並進行定制化加工，消耗大量的時間和精力。為解決這一問題，本研究借用LLM的語境理解與文本生成能力，構建了易於操作的 *ChineseRealTalk* 平台。該平台提供三大核心功能：根據學習者的HSK等級重寫原文的“分級文本轉換工具”；為生詞自動添加拼音和例句的“新詞與難點表達解析工具”；以及

* 덕성여자대학교 중어중문학전공·소프트웨어전공, 부교수

自動生成選擇題與討論主題的“閱讀理解題生成工具”。研究表明，該平台能夠大幅減少教師備課時間。然而，研究同時也發現LLM仍存在若幹局限，包括幻覺現象、文本文化語境的流失以及因知識截止所造成的最新語言現象反映滯後等問題。綜上，LLM在教學資源開發中可作為生成初稿的有效工具。為充分發揮其優勢，教師需要具備主動策展能力，對生成結果進行批判性審查，並通過精細化的提示詞設計提升生成質量。

關鍵詞：大語言模型(LLM)，社交媒體文本，網絡新詞，漢語教學，定制化文本，ChineseRealTalk

I. 머리말

중국어 교육 현장에서 오랫동안 지적되어 온 문제 중 하나는 바로 교실에서 배우는 중국어와 실제 원어민이 구사하는 일상 언어 사이에 간극이 존재한다는 것이다. 교재에 수록된 대화문은 일정한 어휘 수준을 유지하는 문법적인 구문으로 이루어져 있지만, 정작 학습자가 실생활에서 맞닥뜨리는 중국어, 특히 특히 微博(웨이보), 小红书(샤오홍수), 微信(위챗) 등 소셜 미디어 플랫폼의 텍스트는 이와는 전혀 다른 어휘와 문체를 보인다. 가령, ‘今天真的emo了(오늘 정말 우울해)’나 ‘这个绝绝子(이거 완전 최고야)’, ‘躺平就完事了(그냥 드러눕는 거야)’와 같은 표현들은 현대 중국어 사용자들 사이에서 일상적으로 통용되지만, 대다수의 중국어 교재에서는 찾아볼 수 없다. 그렇기에 수년간 중국어를 학습한 중급 이상의 학생조차 중국 드라마나 소셜미디어 텍스트를 처음 접하면 당혹감을 느끼기 마련이다.

이처럼 독특한 소셜미디어 문체는 현재 학생 세대의 감수성과 긴밀하게 연결된다는 점에서 교육의 필요성이 대두된다. 이른바 90后, 00后로 대변되는 중국의 MZ세대는 디지털 환경에서 성장하여 직접 통화보다 텍스트

기반 소통을 편안하게 느끼고, 조직보다 개인의 행복을 중시하는 가치관을 지닌 세대이다. 躺平(tǎngpíng, 드러눕다)이나 内卷(nèijuǎn, 안으로 말리다)과 같은 신조어는 청년 세대가 경험하는 취업난, 고물가, 경쟁 과열이라는 사회적 압박이 표출된 것으로, 한국의 N포세대나 일본의 사토리 세대와 같이 특정한 세대(cohort) 간 유사한 맥락을 공유하고 있다. 이외에도 직장 생활과 청년 문화를 반영하는 打工人(dǎgōngrén, 직장인·월급쟁이), 干饭人(gānfǎnrén, 먹는 데 진심인 사람), 工具人(gōngjùrén, 호구처럼 이용당하는 사람) 등의 신조어는 ‘-하는 사람(人)’이라는 결합 구조를 통해 젊은 세대의 사회 일상과 자조(自嘲)적인 유머를 담아낸다. 이와 같은 인터넷과 소셜미디어에서 유행하는 신조어 학습은 단순한 어휘 학습의 차원을 넘어, 학생들에게 중국의 또래 세대와 중국 사회의 이면을 통찰하는 중요한 교재가 될 수 있다고 생각한다.

이러한 관점에서 송지현(2002), 陸丹妮(2015), 梁雲雲(2021), 鄭金芝(2022), 박민준(2020, 2024), 주장훈(2023), 하두진·박민준(2025) 등 소셜미디어 텍스트를 교육 자료로 활용하려는 시도가 없었던 것은 아니다. 그러나 기존 방식은 교사가 일일이 텍스트를 선별하고, 어려운 표현에 주석을 달고 이해도 확인 문항까지 수작업으로 만들어야 하는 한계가 있었다. 필자를 비롯하여 대다수 교사는 시의성을 지닌 온라인 텍스트 교육의 필요성을 인식하고 있다. 하지만 온라인 자료를 수업에 활용하기 위해서는 상당한 시간과 노력을 들여야 하며, 이는 전적으로 교사 개인의 부담으로 이어진다. 이런 문제로 인해, 교육 현장에서 온라인 소셜미디어 텍스트의 활용은 여전히 제한적으로 이루어지고 있는 실정이다.

이러한 상황에서 2022년 이후 ChatGPT로 대표되는 대형언어모델(Large Language Model, 이하 LLM)의 급격한 발전은 이 오래된 문제에 대한 새로운 해법을 제시하고 있다. LLM은 방대한 언어 데이터로 사전 학습된 결과, 텍스트의 맥락을 이해하고 지정된 조건에 맞게 재작성하는 능력을 갖추고 있다. 즉, 소셜미디어 원문을 입력하면 HSK(汉语水平考试) 수준에 맞게 난이도를 조정하거나, 신조어에 병음과 의미 주석을 자동으로 생성하

거나, 읽기 후 활동 문항을 생성하는 작업이 가능해진 것이다.

이에 본고는 LLM을 활용하여 소셜미디어 중국어 텍스트를 교육 자료로 변환하는 방법론을 탐구하고, 이를 구현한 인터랙티브 플랫폼 Chinese RealTalk를 구축함으로써 LLM 시대 중국어 교육의 새로운 가능성과 한계를 함께 논의하고자 한다. 논문의 구성은 다음과 같다. 2장에서는 소셜미디어 중국어의 언어적 특성과 교육 자료 활용에 있어서의 어려운 점들을 살피고, 3장에서는 ChineseRealTalk 플랫폼의 구현 및 기능을 소개하고, 4장에서 교육 활용 사례를 분석한 후, 5장에서 결론 및 향후 과제를 제시할 것이다.

II. 소셜미디어 텍스트의 특성과 교육 활용도

1. 신조어의 언어적 특징

최근 중국의 소셜미디어 플랫폼은 끊임없이 생성되고 소멸하는 인터넷 신조어(网络新词)를 중심으로 독자적인 언어 생태계를 형성하고 있다. 이러한 신조어가 만들어지는 기제는 매우 다양한데, 크게 네 가지 유형으로 나누어 볼 수 있다.

첫째, 의미 전용 및 은유적 확장이다. 이는 기존 어휘에 은유를 통해 새로운 뜻을 부여하거나 그 쓰임을 확장하는 경우이다. 가령 躺平(드러눅다)이 경쟁을 포기하고 최소한의 삶을 선택하는 태도를 뜻하게 된 것이나, 内卷(안으로 말리다)이 과도한 내부 경쟁과 노력의 무의미함을 표현하는 신조어로 정착된 것이 대표적이다. 또한, 裂开(쪼개지다)가 황당하거나 당혹스러운 심경을 표현하고, 破防(방어선이 뚫리다)이 ‘정신줄을 놓다’라는 뜻으로 사용되는 등 은유적 확장을 거친 어휘들도 이에 속한다. 이밖에도 给力(방언. ‘힘을 내라’ 혹은 ‘힘이 난다’)가 ‘최고다, 대박이다’라는 의미로,

社死(사회적 죽음)가 ‘공개적 망신, 이미지 폭망’의 뜻으로 확장되거나, 프랑스 베르사유 궁전의 화려한 이미지를 은유적으로 차용하여 겉으로는 검소한 척하면서 은근히 사랑하는 행위를 가리키는 凡尔赛(Fán'èrsài) 역시 기존 어휘의 새로운 활용 양상을 잘 보여주고 있다.

둘째, 외래어, 특히 영어 단어를 발음이나 알파벳 그대로 차용하거나 의역하여 활용하는 표현이다. emo(우울하다), diss(비난하다), PUA(Pick-Up Artist, 被PUA: 가스라이팅), 硬核(hardcore) 등은 젊은 세대의 소통에서 일상적으로 쓰이고 있다.

셋째, 해음어(谐音词) 및 축약어(缩略语)이다. yyds(永远的神, 영원한 신)는 최고라는 뜻의 극찬으로 널리 쓰이며, 绝绝子(jué jué zi)는 绝了(끝내준다)를 강조하는 과생어로 긍정적, 부정적 상황 모두에 쓰인다. 발음이 유사한 글자를 활용한 해음어로는 冲啊(돌격!) 대신 쓰는 冲鸭(가즈아!), 照片(사진)과 骗(속이다)을 합성하여 과도한 보정 사진을 꼬집는 照骗(셀기꾼) 등이 있다.

넷째, 특정 사회적 역할 및 군상을 지칭하는 결합이다. 앞서 언급한 打工仔, 干饭人, 工具人과 같은 ‘N+人’ 결합 구조 외에도, 분위기를 띄우는 气氛组(분위기 메이커), 위험을 무릅쓰고 현장을 지키는 逆行者(역행자) 등 인물과 집단을 나타내는 형태소가 생산적으로 결합하고 있다. 이는 특정 사회적 역할을 지칭하는 생산적인 조어 방식인 동시에, 그에 대한 청년층의 인식과 태도를 은연중에 드러낸다.

〈표 1〉 중국 소셜미디어 주요 신조어 현황(2018~2022)

연도	신조어	성격	유형	한국어 의미
2018	佛系	中性	의미전용	집착 없이 여유롭게 사는 태도 (불교적 무심)
2018	杠精	中性	역할군상	무조건 반박하는 사람 (유사어: ‘트롤’)
2018	真香	中性	은유확장	‘진짜 향기롭다’ → 처음에 거부하다가 결국 좋아함
2018	C位	中性	외래어	‘center position’ → 가장 주목받는 위치

연도	신조어	성격	유형	한국어 의미
2019	柠檬精	中性	은유확장	‘레몬 요정’ → 질투심 많은 사람
2019	硬核	正向	외래어	‘hardcore’ → 본격적인, 수준 높은
2019	奥利给	正向	은유확장	‘힘내라’ ‘파이팅’의 강조 표현
2019	上头	中性	의미전용	머리로 올라오다 → 빠져들다 열중하다
2020	内卷	负向	의미전용	과도한 경쟁으로 인한 소모전 쓸데없는 노력 경쟁
2020	打工人	中性	역할군상	임금노동자 직장인 (자조적 표현)
2020	凡尔赛	中性	은유확장	둘러서 자랑하기, 우아한 척 자랑하는 행위
2020	爷青回	中性	축약어	‘내 청춘이 돌아왔다’ (노스탤지어적 표현)
2020	集美	中性	해음어	‘姐妹(jiěmèi 자매)’의 해음어
2020	带货	中性	의미전용	상품을 홍보·판매하다 (라이브커머스 용어)
2021	YYDS	中性	축약어	‘永远的神(영원한 신)’ → 최고야 짱이야
2021	绝绝子	中性	파생어	‘絶了’의 강조형 → 완전 최고야 대박이야
2021	躺平	中性	은유확장	‘드러눅다’ → 경쟁 포기, 최소한의 삶 선택
2021	破防	中性	은유확장	‘방어가 뚫리다’ → 마음이 흔들리다, 감동받다
2021	emo	中性	외래어	emotional → 우울하다, 감성에 빠지다
2021	元宇宙	中性	외래어	metaverse → 메타버스, 가상세계
2021	冲鸭	正向	해음어	‘冲啊(돌격!)’의 변형 → 가즈아! 파이팅!
2021	气氛组	中性	역할군상	‘분위기 조성 그룹’ → 분위기 메이커, 인싸
2021	逆行者	正向	역할군상	‘거슬러 가는 자’ → 위험에도 현장을 지키는 사람
2021	社死	负向	의미전용	‘사회적 죽음’ → 공개적 망신, 폭망
2021	工具人	负向	역할군상	‘도구 같은 사람’ → 이용당하는 사람, 호구
2021	干饭人	中性	역할군상	‘밥 먹는 사람’ → 먹는 데 진심인 사람
2021	给力	正向	의미전용	‘힘내라/힘이 난다’ → 최고다, 대박이다
2022	照骗	中性	해음어	‘照片(사진)+骗(속임)’ 합성 → 셀기꾼 필터 과도 사용

연도	신조어	성격	유형	한국어 의미
2022	累丑	中性	축약어	‘疲累(피로)+丑(추함)’ → 피곤해서 초췌해지다
2022	慢就业	中性	역할군상	‘느린 취업’ → 졸업 후 바로 취업을 하지 않고 유보하는 현상
2022	全职儿女	中性	역할군상	‘전업 자녀’ → 부모 집에서 가사 분담을 직업으로 삼는 청년
2022	电子榨菜	中性	은유확장	‘전자 반찬’ → 식사 중에 보는 영상 콘텐츠

2. 발생적 특징과 교육적 과제

소셜미디어 텍스트의 이러한 입체적이고 역동적인 속성은 인터넷 플랫폼 환경과 밀접한 관련이 있다. 웨이보(微博)와 같은 소셜미디어는 모바일 환경에 최적화되어 있어, 길고 정제된 문장보다는 짧고 즉각적인 단문 중심의 소통이 발달하였다. 100자 안팎의 짧은 텍스트만을 허용하기 때문에 사진, 동영상, 이모티콘 같은 시각 자료와 결합하여 의미를 보완하는 경향이 나타난다. 또한 소셜미디어 텍스트는 개별 문장 하나만으로 의미가 완성되기보다는, 게시물 간 연결 구조 속에서 의미가 형성되는 담화적 특성을 지닌다. 즉, 특정 게시물이 관심을 받으면 그 thread를 중심으로 새로운 교류가 시작되고, 해시태그를 통해 더 넓은 사용자 집단으로 확산되는 방식으로 작동한다(박민준, 2020). 이에 더해 고정된 문법 규범에서 벗어난 생략 구조나 영어와 중국어가 혼용되는 코드스위칭(예: 今天去咖啡馆 check in了)도 빈번하게 나타난다. 결국 소셜미디어 텍스트는 정적인 문장들의 집합이라기보다, 이동성과 전파성이 강한 동적인 흐름으로 이해할 필요가 있다.

이러한 소셜미디어 텍스트의 즉시성과 트렌디함은 디지털 환경에 익숙한 MZ세대 학생들의 흥미를 끌기에 충분하지만, 역설적으로 비정형적이고 맥락 의존적인 속성 때문에 일반적인 교재보다 해석하고 교육용으로

가공하기가 훨씬 까다롭다. 躺平이나 内卷이 단순한 동사가 아니라 현대 중국 청년 세대의 집단적 정서와 사회 비판 의식을 드러내듯이, 신조어는 단순한 유행어를 넘어 특정 세대나 집단의 정서를 공유하는 표지(marker)로 기능하며, 때로는 온라인 혐오 표현(hate speech)처럼 특정 집단, 성별, 인종에 대한 편견을 노출하기도 한다(朴敏浚, 2025). 따라서 이를 교실로 가져오기 위해서는 단순히 어려운 표현을 쉬운 말로 바꾸는 것을 넘어, 교육적으로 어떤 텍스트를 어떻게 가르칠 것인지에 대한 교사의 세심한 교육적 판단이 요구된다.

3. 교육 자료로서의 가능성과 도전

소셜미디어 텍스트를 교육에 활용하는 것은 분명한 장점이 있다. 학생들에게 실제 원어민의 언어 사용 맥락을 생생하게 보여줌으로써 호기심과 학습 동기를 촉진하고, 동시대 중국 사회와 문화에 대한 이해도를 동시에 높일 수 있기 때문이다.

그러나 교육 현장에 바로 적용하기에는 몇 가지 현실적인 어려움 또한 존재한다. 가장 큰 문제는 소셜미디어 텍스트에 섞여 있는 유해 콘텐츠다. 혐오 발언이나 성차별적 표현 등을 걸러내고 교육적으로 유의미한 텍스트만 선별하는 작업은 교사에게 상당한 부담이 된다. 또한 신조어의 빠른 변화 주기로 인해 오늘의 최신 표현이 내일은 이미 구식이 될 수 있기 때문에 교사가 지속적으로 최신 언어 변화를 추적해야 하는 수고가 따른다. 따라서 기존의 전통적인 교재 집필 방식으로는 이러한 역동적인 언어 변화를 따라가기가 어려운데, 梁雲雲(2021) 등의 선행 연구에서 인터넷 유행어를 어휘 교육에 접목하려는 시도가 있었으나 텍스트 전체를 학생 수준에 맞게 변환하고 체계적으로 주석을 다는 구체적인 방법론까지 나아가지는 못했다.

본고는 바로 이러한 한계점을 극복하는 데 대형언어모델(LLM)이 매우 유효하고 강력한 보조 도구가 될 수 있음을 증명할 것이다. 朴敏浚(2016)이 중국어 정보 처리 플랫폼을 초급 학습자를 위해 개발한 이래, 중국어

교육에서 디지털 도구 활용에 대한 관심이 꾸준히 증가해 왔으나, LLM의 등장 이후 이 논의는 완전히 새로운 차원으로 발전하고 있다. 다음 장에서는 LLM의 중국어 처리 능력을 바탕으로, 이를 실제 교육 자료 제작에 어떻게 적용할 수 있는지 구체적인 세 가지 기술을 살펴보고자 한다.

III. ChineseRealTalk 인터랙티브 교육 플랫폼 구현

1. LLM의 중국어 처리 능력

최근 대형언어모델(LLM)은 수천억 개의 파라미터와 방대한 다국어 코퍼스를 학습한 결과, 중국어 텍스트에 대해 놀라운 수준의 이해 및 생성 능력을 보여주고 있다. GPT(OpenAI), Claude(Anthropic), DeepSeek(深度求索) 등 주요 LLM들은 표준 중국어는 물론이고 끊임없이 변하는 인터넷 신조어나 방언, 심지어 영어와 중국어가 섞인 코드스위칭 문장까지 상당한 수준으로 처리해 낸다.

이러한 LLM의 자연어 처리 능력은 중국어 교육 측면에서 크게 두 가지 긍정적인 가능성을 시사한다. 첫째는 텍스트의 숨은 맥락을 파악하는 이해 능력이다. 가령 内卷이나 逆行者처럼 특수한 사회적 맥락이 존재하는 표현도, 적절한 프롬프트를 입력하면 모델이 그 뉘앙스를 파악하여 학생이 이해하기 쉬운 한국어나 표준 중국어로 풀어서 설명해 준다. 둘째는 교사의 의도에 맞게 텍스트를 재구성하는 생성(generation) 능력이다. 교사가 특정 HSK 급수를 지정하면, 단순히 어려운 단어를 쉬운 단어로 바꾸는 수준을 넘어 문장 구조 전체를 해당 난이도에 맞게 다시 써준다. 특히 주목할 만한 점은 ‘HSK 4급 수준으로 바꾸되 원문의 감성적인 분위기는 최대한 살려라’라는 식의 복잡한 요청에도 이를 충실히 따르는 지시 수행 능력이 탁월하다는 것이다. Park&Lee(2025)에서 검색증강생성(RAG)를 활

용해 언어학 지식을 전수함으로써 LLM의 사실성 추론 능력을 높였듯이, LLM이 중국어 교재 데이터베이스와 연동한다면 더욱 정밀한 교육용 텍스트 변환 및 생성도 가능할 것이며, 이는 물론 교사의 소셜미디어 텍스트 교재화 가공 부담을 획기적으로 덜어줄 것이다.

2. LLM 기반 인터랙티브 교육 플랫폼의 주요 기능

본고의 소셜미디어 텍스트 교재화 플랫폼 ChineseRealTalk는 프로그래밍 지식이 없는 중국어 교사와 학생 누구나 LLM의 텍스트 변환 기술을 손쉽게 활용할 수 있도록 개발된 인터랙티브 플랫폼이다.

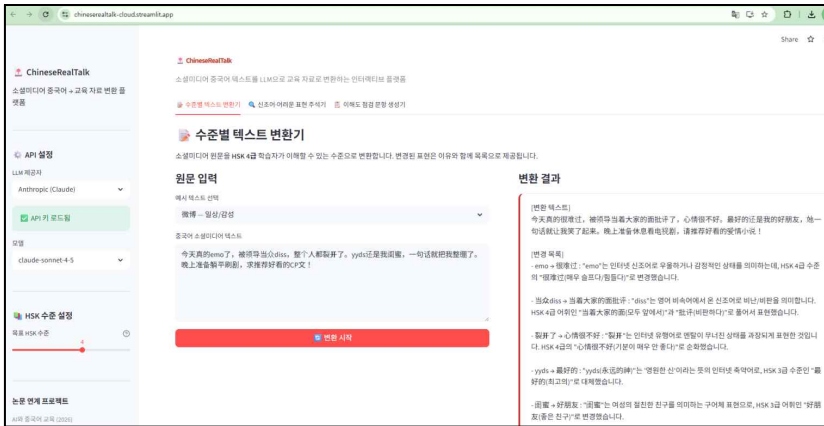
이 플랫폼은 세 가지 핵심 원칙을 바탕으로 설계되었다. 첫째, 누구나 쉽게 접근할 수 있는 접근성이다. Streamlit¹⁾ 기반의 웹 브라우저(<https://chineserealtalk.streamlit.app/>) 환경으로 구축되어 별도의 프로그램 설치 없이 바로 사용할 수 있다²⁾. 둘째, 교사의 재량을 극대화하는 유연성이다. 사용 환경에 따라 Claude와 GPT 모델 중 하나를 선택할 수 있으며, 학생들의 수준에 맞추어 목표 HSK 등급을 1급부터 6급까지 슬라이더로 자유롭게 조절할 수 있다. 셋째, 지속적인 발전을 위한 재현 가능성이다. 플랫폼의 전체 소스 코드를 GitHub(<https://github.com/karmalet-DS/ChineseRealTalk>)에 공개하여, 다른 연구자나 일선 교사들이 각자의 필요에 맞게 기능을 수정하고 확장할 수 있도록 하였다.

본 플랫폼은 크게 세 가지 주요 탭(기능)으로 구성되어 있으며, 각각의 기능은 다음과 같다.

-
- 1) Streamlit은 웹 개발 지식이 없는 연구자도 파이썬(Python) 언어만으로 AI모델이나 데이터 분석 결과를 쉽게 구현하고 배포할 수 있도록 지원하는 오픈소스 도구이다.
 - 2) 2026년 4월까지 한시적으로 무료 운영할 계획이며, 이후에는 사용자 개인의 API Key를 입력하여 사용가능하다.

1) 수준별 텍스트 변환기

첫 번째 기능은 소셜미디어 원문을 학습자 수준에 맞게 재작성하는 텍스트 난이도 조정(Text Leveling) 기능이다. 사용자가 微博나 小红书的 텍스트를 직접 붙여넣거나 혹은 예시 텍스트를 대상으로 원하는 HSK 급수를 설정하면 텍스트 변환이 시작된다. 이 때 입력한 글은 사이드바에서 설정한 HSK 수준으로 변환이 이루어지며, 소셜미디어 특유의 과도한 생략 구조는 규범적인 문장으로 복원된다. 화면에는 변환된 전체 텍스트와 함께 어떤 표현이 왜 바뀌었는지를 보여주는 변경 목록이 나란히 표시되며(그림 1), 이 결과물은 수업 자료로 쓰기 좋게 텍스트(.txt) 파일로 바로 다운로드할 수 있다.



〈그림 1〉 수준별 텍스트 변환기

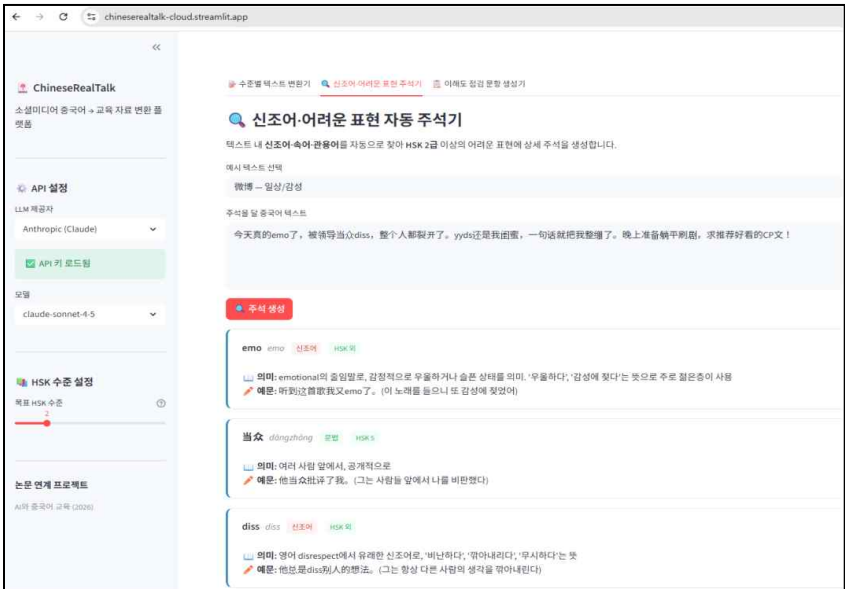
이를 구현하기 위해 프롬프트에는 LLM에 ‘중국어 교육 전문가’라는 역할을 명확히 부여하고(Role Prompting), 핵심 의미 보존과 HSK 수준 적용이라는 구체적인 규칙을 지시하였으며, 교사가 쉽게 검토할 수 있도록 변경 전·후의 이유를 체계적으로 출력하도록 설계하였다.

〈표 2〉 텍스트 교육화 변환기의 시스템 프롬프트 및 수행 결과

<p>[시스템 프롬프트]</p> <p>당신은 중국어 교육 전문가입니다. 소셜미디어에서 가져온 중국어 텍스트를 지정된 HSK 수준의 학습자가 이해할 수 있도록 변환합니다. 규칙: 1. 원문의 핵심 의미와 문화적 맥락을 보존하세요. 2. 신조어·속어·어려운 표현은 HSK 수준에 맞는 표현으로 바꾸세요. 3. 변환된 텍스트와 함께, 변경된 표현 목록을 제공하세요: - 원문 표현 → 변환 표현 : 이유</p>
<p>[사용자 프롬프트]</p> <p>다음 중국어 소셜미디어 텍스트를 HSK {수준}급 학습자 수준으로 변환해 주세요. [원문]: {텍스트}</p>
<p>[원문 Bilibili]</p> <p>这才是优秀的节奏型作品，音乐适合，节奏精准，调音优秀，最重要的是有故事有剧情，比硬踩点的作品舒适多了，这次一定，这就是TMD惊喜，希望up继续努力，产出更多优秀作品</p>
<p>[HSK2급 변환 텍스트]</p> <p>这个作品很好。音乐很好，节奏也很好。最重要的是有故事。比别的作品好看多了。这次真的很好，这是一个惊喜。希望你继续努力，做更多好作品。</p>
<p>[변경 목록]</p> <p>- 这才是优秀的节奏型作品 → 这个作品很好: “节奏型作品”은 전문 용어 이므로 HSK 2급의 기본 형용사 “好”로 단순화 - 音乐适合 → 音乐很好: “适合”(적합하다)는 HSK 3급 어휘이므로 HSK 2급의 “很好”로 변경 - 节奏精准 → 节奏也很好: “精准”(정확하다)은 고급 어휘이므로 “很好”로 단순화 - 调音优秀 → (삭제): “调音”(믹싱)은 전문 용어이며 HSK 2급 수준을 넘어 서므로 생략 - 有故事有剧情 → 有故事: “剧情”(줄거리)은 HSK 4급 어휘이므로 “故事”만 사용</p>

- 硬踩点的作品 → 别的作品: “硬踩点”(비트에 강제로 맞추기)은 전문 용어이므로 “别的”(다른)로 단순화
- 舒适多了 → 好看多了: “舒适”(편안하다)은 HSK 3급 이상이므로 HSK 2급의 “好看”으로 변경
- 这次一定 → 这次真的很好: 인터넷 유행어 “这次一定”을 완전한 문장으로 변경
- 这就是TMD惊喜 → 这是一个惊喜: “TMD”는 속어이므로 삭제하고 문장 구조 단순화
- up → 你: “up”(업로더)는 인터넷 신조어이므로 기본 대명사 “你”로 변경
- 产出 → 做: “产出”(산출하다)은 HSK 4급 어휘이므로 HSK 1급의 “做”로 변경

2) 신조어·어려운 표현 주석기



〈그림 2〉 신조어·어려운 표현 주석기

두 번째 기능은 텍스트 내의 어려운 표현을 자동으로 찾아내어 학습용

주석을 달아주는 기능(Auto-Glossing)이다. 텍스트를 입력하면 LLM이 학생들이 이해하기 어려운 법한 단어들을 감지하여 각각의 단어마다 카드 형태의 주석을 생성해 준다. 이 카드에는 해당 표현의 병음, 단어의 유형(신조어·속어·관용어 등), 한국어 의미, 실제 쓰임을 보여주는 중국어 예문, 그리고 해당하는 HSK 등급과 같은 세부 정보가 기록되며, 이렇게 생성된 전체 주석 목록은 csv 파일로 저장할 수 있어 Anki나 Quizlet 등의 플래시카드 기반의 단어 학습 프로그램에 간편하게 연동하고 활용하기 쉽도록 설계하였다.

〈표 3〉 신조어·어려운 표현 주석기의 시스템 프롬프트

```
[시스템 프롬프트]
당신은 중국어 언어학 전문가이자 교사입니다.
중국어 텍스트에서 학습자에게 어려울 수 있는 표현을 찾아
상세한 주석을 달아줍니다.

다음 형식으로 JSON 배열을 반환하세요:
[
  {
    "표현": "원문 표현",
    "병음": "pinyin",
    "유형": "신조어|속어|관용어|문법",
    "의미": "한국어 설명",
    "예문": "다른 예문 (중국어)",
    "HSK등급": "해당 없음 또는 1~6"
  }
]
```

3) 이해도 점검 문항 생성기

마지막은 텍스트 학습을 마치고 학습 내용 점검용으로 활용할 수 있는 다양한 퀴즈를 자동 생성하는 기능이다. 교사가 객관식, 단답형, O/X, 서술형 등 원하는 문항의 종류와 개수를 선택하면, 본문 내용에 기초한 이해도 확인 문제, 학생들과 토의할 수 있는 토론 주제, 그리고 본문의 핵심

어휘 리스트가 바로 만들어진다. 생성된 퀴즈 문항 역시 텍스트 파일로 다운로드하여 실제 교실이나 온라인 수업에서 바로 배포할 수 있다.

〈표 4〉 이해도 점검 문항 생성기 실행 결과

HSK 5급 학습자용 읽기 후 활동

[이해 확인 문항]

Q1. 이 글에서 작품에 대한 평가로 옳지 않은 것은?

- ① 음악이 적합하다
- ② 리듬이 정확하다
- ③ 스토리와 줄거리가 있다
- ④ 억지로 박자를 맞춘 작품이다

정답: ④

해설: 글쓴이는 “比硬踩点的作品舒适多了(억지로 박자를 맞춘 작품보다 훨씬 편안하다)”라고 말하며, 이 작품은 억지로 박자를 맞춘 작품이 아니라고 평가하고 있습니다.

Q2. 글쓴이가 제작자(up)에게 바라는 것은 무엇입니까?

정답: 계속 노력해서 더 많은 우수한 작품을 만들어내는 것 / 더 우수한 작품을 제작하는 것

해설: 마지막 부분 “希望up继续努力, 产出更多优秀作品”에서 제작자가 계속 노력하여 더 많은 우수한 작품을 만들어내길 바란다고 명확히 표현하고 있습니다.

[토론 질문]

1. 여러분은 영상이나 음악 작품을 평가할 때 어떤 요소를 가장 중요하게 생각하나요? (예: 기술적 완성도, 스토리, 감정 전달 등) 그 이유는 무엇인가요?

2. “硬踩点(억지로 박자를 맞추)”처럼 기술적으로는 정확하지만 자연스럽지 않은 작품과, 약간 부족하더라도 감정이 풍부한 작품 중 어느 것을 더 선호하나요? 왜 그렇게 생각하나요?

[핵심 어휘 정리]

- 节奏: 리듬, 박자 (jiézōu)
- 例: 这首歌的节奏很快。(이 노래의 리듬이 아주 빠르다.)

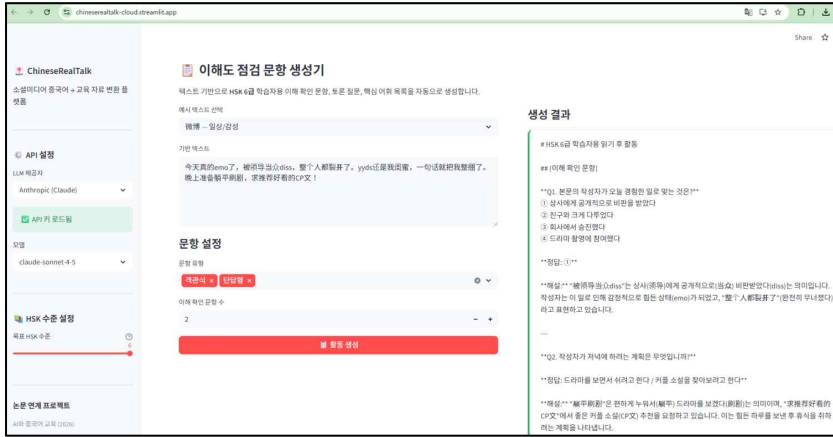
- 精准: 정확하다, 정밀하다 (jīngzhǔn)
- 例: 他的分析非常精准。(그의 분석은 매우 정확하다.)

- 剧情: 줄거리, 스토리 (jùqíng)
- 例: 这部电影的剧情很吸引人。(이 영화의 줄거리는 매우 흥미롭다.)

- 产出: 생산하다, 만들어내다 (chǎnchū)
- 例: 这个工厂每天产出大量产品。(이 공장은 매일 대량의 제품을 생산한다.)

- 舒适: 편안하다, 쾌적하다 (shūshi)
- 例: 这个沙发坐着很舒适。(이 소파는 앉으면 매우 편안하다.)

이처럼 ChineseRealTalk 플랫폼을 사용하면 소셜미디어 텍스트 한 토막을 가지고 위에서 살펴본 세 가지 기능을 통해 클릭 한 번만으로도 맞춤형 교육 자료를 만들 수 있다. 기존에 교사가 일일이 사전과 씨름하며 수작업으로 단어장을 만들고 수업 자료를 준비하던 방식에 비하면 그야말로 획기적인 효율성 향상이라 하지 않을 수 없다. 다음 장에서는 ChineseRealTalk 플랫폼의 수업 활용 예시를 소개한다.



〈그림 3〉 이해도 점검 문항 생성기

IV. 논의 및 활용 사례 탐구

1. 수업 적용 시나리오

위에서 소개한 ChineseRealTalk 플랫폼이 실제 수업 준비 과정에 어떻게 적용될 수 있을까? 교사는 수업 전 微博나 小红书에서 당일 학습 주제와 맞닿아 있는 게시글을 하나 선정한다. 가령 ‘취업과 직장 생활’이 주제라면 打工工人 관련 밈(meme)이나 内卷 현상을 다룬 단문을 고른다.

먼저 해당 텍스트를 플랫폼의 3가지 기능 탭에 입력하고 학생들의 수준에 맞춰 HSK 급수를 선택하면, 단시간 내에 난이도가 조정된 텍스트, 핵심 어휘 단어장 및 문화적 맥락을 반영한 토론 주제(예: 躺平 현상은 개인의 선택인가, 사회 구조의 문제인가?)가 화면에 출력되며, ‘다운로드’ 버튼을 클릭하면 파일로 저장할 수도 있다. 수업 준비 단계에서 교사는 이 초안을 검토하여 필요한 부분을 선정하여 유인물로 제작하여 학습관리시

스텝(LMS)에 업로드한다. 수업 초반에는 ‘신조어·어려운 표현 주석기’를 통해 어휘 카드를 생성하여 수업 전 단어 암기나 워밍업 활동에 활용하고, 수업 말미에는 ‘이해도 점검 문항 생성기’가 만든 학습 점검용 객관식 퀴즈와 단답식 문항을 바탕으로 조별 시합이나 소그룹 토론 등의 읽기 후 활동을 진행할 수 있을 것이다. 이러한 과정을 거치는 데 소요되는 수업 준비 시간은 교사의 검토 시간을 포함해도 15~20분 남짓으로, 기존의 수작업 방식과 비교할 때 상당한 시간 절감 효과를 기대할 수 있다.

2. 텍스트 자동 해석 예시

ChineseRealTalk 플랫폼에서 실제 소셜미디어 텍스트가 어떻게 자동 해석되는지 다음의 微博 예문을 통해 살펴보자.

[원문] 今天真的emo了，被领导当众diss，整个人都裂开了。yyds还是我闺蜜，一句话就把我整绷了。晚上准备躺平刷刷，求推荐好看的CP文！

이 짧은 텍스트 안에는 ‘emo(우울하다)’, ‘diss(비난하다)’, ‘裂开了(정신이 나갔다)’, ‘yyds(최고야)’, ‘整绷了(웃겨서 기분이 풀렸다)’, ‘躺平(그냥 드러눕다)’, ‘CP文(커플물 팬픽션)’ 등 정규 교재에서는 전혀 다루지 않는 신조어와 유행어가 마구 섞여 있다. 문화적 맥락을 모르는 학생들을 위해 HSK 4급 수준으로 변환한 결과는 다음과 같다.

[변환] 今天真的很难过，被领导在大家面前批评，感到非常沮丧。我的好朋友是最好的，她说了一句话就让我开心起来了。晚上打算放松一下，看看电视剧，请推荐好看的爱情故事！

〈표 5〉 텍스트 어휘 해석 예시 (GPT)

원문 표현	HSK 4급 변환	변환 이유
emo了	很难过	음역 신조어 → HSK 4급 감정 어휘
diss	批评	영어 차용어 → 표준 동사
裂开了	cc	은유적 신체 표현 → 명시적 감정 표현
yyds	最好的	이니셜 약어 → 형용사구
整绷了	让我开心起来	인터넷 속어 → 서술어절
躺平	放松一下	사회적 신조어 → 일반 동사구
CP文	爱情故事	팬덤 용어 → 일반 명사

변환 결과를 살펴보면 외래어인 ‘emo’는 HSK 4급 감정 어휘인 ‘很难过’로, ‘diss’는 중국어 동사 ‘批评’으로 바뀌었다. 은유적 표현인 ‘裂开了’는 명시적 감정 표현인 ‘感到非常沮丧’으로, 인터넷 속어 ‘整绷了’라는 표현은 ‘让我开心起来’로 직접적으로 풀이되었다. 사회적 맥락 의미가 강한 ‘躺平’은 보다 일반적인 의미의 ‘放松一下’로, 팬덤 용어인 ‘CP文’은 일반 명사인 ‘爱情故事’로 바뀐 것을 확인할 수 있다. 또한 동일한 원문을 HSK 6급 수준으로 변환할 경우에는 ‘emo’가 ‘情绪低落’으로, ‘diss’가 ‘公开指责’로, ‘裂开了’가 ‘崩溃至极’로 바뀌어 좀 더 수준 높은 서면체 어휘가 구사됨을 확인하였다.

3. LLM 활용 소셜미디어 텍스트 교재화의 한계 및 보완점

신조어 주석 생성 기능을 분석한 결과, emo나 yyds 같은 단순 축약어는 비교적 안정적으로 처리되었으나, 최근(2024~2025년) 유행하는 신조어에 대해서는 생성 결과가 불안정했다. 무엇보다 주의해야 할 점은 LLM 특유의 환각(Hallucination) 현상이다. 예를 들어, 搞钱(돈 버는 데 몰두하다)이라는 단어의 어원에 대해 LLM은 ‘改钱에서 유래하여 돈을 바꾸다는 뜻에서 의미가 변화했다’고 그럴싸하게 설명하였으나 이는 사실과 다른데, 搞

钱은 단순히 搞(하다)와 钱(돈)이 습관적으로 결합하여 사용된 관용 표현 일 뿐이다. 또한, 破防(방어선이 뚫리다, 마음이 흔들리다)에 대해서도 ‘게임 용어에서 유래하여 상대방의 방어를 뚫는다는 뜻에서 감정의 붕괴를 표현하게 되었다’고 설명하였는데, 어느 정도 개연성은 있으나 이 또한 학술적으로 명확히 검증된 어원은 아니다. 이처럼 모델이 생성한 이유 중에 근거가 부족한 정보가 학생들에게 그대로 전달될 경우 잘못된 언어 지식을 심어줄 위험이 크다. 즉 LLM을 교육 현장에 활용할 때에는 교사의 비판적이고 꼼꼼한 사전 검토가 절대적으로 필요하다.

이처럼 LLM을 활용한 소셜미디어 텍스트의 교재(教材)화는 분명히 잠재력을 가지고 있지만, 동시에 몇 가지 뚜렷한 한계점도 드러내고 있다.

첫째, 문화적 맥락의 손실이다. 앞서 <표 4>의 예시처럼 躺平을 단순히 放松(휴식하다)으로 치환하면 의미 전달은 되지만, 그 안에 내재된 현대 중국 청년들의 치열한 피로감과 사회적인 탈경쟁 분위기는 완전히 사라져 버린다. 따라서 텍스트가 담고 있는 사회문화적 배경은 교사가 직접 보충하여 설명해야 한다.

둘째, 최신 언어 변화에 대한 지연(遲延) 현상이다. LLM은 특정 시점까지의 데이터로 학습을 마치는 지식 컷오프(Knowledge Cutoff)의 제약을 받으므로, 그 이후에 새롭게 생겨난 조어 현상에는 정확도가 떨어지게 된다. 가령 본 연구의 ChineseRealTalk 플랫폼이 탑재하고 있는 GPT-4o 모델의 지식 컷오프는 2023년 10월에 머물러 있으며, 2025년 하반기에 출시된 Anthropic의 Claude 4.5 제품군(Opus, Sonnet, Haiku) 역시 컷오프 날짜가 2025년 7월로 제한되어 있다. 이처럼 LLM의 태생적인 지식의 시차 때문에 모델이 신조어의 맥락을 인식하지 못하거나 엉뚱한 어원을 지어내는 등의 위험성이 상존한다.

셋째, 프롬프트 및 외부 어휘 데이터베이스에 대한 의존성이다. 먼저, 텍스트 변환의 질은 교사의 지시(prompting) 능력에 크게 좌우된다. 단순히 ‘HSK 4급으로 바꿔줘’라고 지시할 때와, ‘원문의 감성을 살리면서 HSK 4급 범위 내의 유의어로 바꾸고 그 이유를 표로 정리해 줘’라고 구체적으

로 지시할 때의 결과물은 질적으로 큰 차이를 보인다. 이는 인공지능을 능숙하게 다루는 프롬프팅 능력이 향후 교사의 새로운 핵심 역량으로 자리잡을 것임을 시사한다. 또한 텍스트 난이도 조정을 위해 참조하는 외부 오픈소스 어휘부(lexicon)의 한계도 존재한다. 중국 중외언어교류협력센터(语合中心, 国家汉办)가 2026년부터 HSK를 총 9등급 체계(HSK 3.0)로 전면 개편함에 따라, 현재 본 플랫폼이 의존하는 깃허브(GitHub)의 오픈소스 어휘부³⁾는 HSK 2.0의 6급 체계(wordlists/old)와 HSK 3.0의 9급 체계(wordlists/new)를 모두 반영하고 있다. 그럼에도 불구하고, 신조어·밈 어휘 등 소셜미디어 특유의 최신 표현은 아직 어느 쪽에도 수록되어 있지 않다. 따라서 최신화되지 않은 기초 어휘부 때문에 변환된 텍스트가 소셜미디어 어휘를 완벽히 반영하지 못하는 점 또한 향후 자체적인 어휘부 업데이트를 통해 해결해야 할 과제로 남아 있다.

V. 맺음말

본 연구는 LLM을 활용하여 소셜미디어 중국어 텍스트를 교육 자료로 변환하는 방법론을 모색하고, 이를 구현한 인터랙티브 교육 플랫폼 Chinese RealTalk를 소개하였다. LLM은 소셜미디어 텍스트의 HSK 수준별 변환, 신조어 주석 생성, 이해도 점검 문항 제작 등의 수업 준비 작업에서 교육적으로 활용 가능한 수준의 성능을 보였으며, 이로 인해 교사들이 과거 일일이 텍스트를 찾아 가공하던 시간과 노력을 대폭 절감할 수 있게 되었다.

이러한 연구 결과를 바탕으로 중국어 교육 현장에 다음과 같은 세 가지 실질적인 제언을 드리고자 한다.

첫째, LLM의 생성 결과는 반드시 교사의 철저한 검토를 거쳐야 한다.

3) Complete HSK Vocabulary
<https://github.com/drkameleon/complete-hsk-vocabulary>

신조어 어원 설명에서 확인된 할루시네이션 사례에서 보듯, LLM 생성 결과를 무비판적으로 수용하는 것은 교육적으로 위험할 수 있다. 교사는 AI를 보조 도구로 활용하되, 최종 검토자로서의 책임을 잃지 말아야 한다.

둘째, 프롬프트를 정교하게 설계하는 능력이 교사의 새로운 필수 자질이 되어야 한다. LLM에게 어떻게 질문하고 조건을 부여하느냐에 따라 생성되는 교육 자료의 질이 크게 달라지는 만큼, 향후 중국어 교사 연수나 교육 과정에 프롬프트 역량 강화가 포함될 필요가 있다.

셋째, ChineseRealTalk를 수업 자료 개발의 출발점으로 활용할 수 있다. 교사가 제로베이스에서 수업자료를 만들기보다, 플랫폼이 생성한 변환 텍스트, 주석, 테스트 문항을 교사가 수정·보완하여 완성도 높은 교육 자료로 발전시키는 방식이 현실적이다. 결국 LLM은 교사의 역할을 완전히 대체하는 만능 도구가 아니라, 교사의 수고를 덜어주는 훌륭한 1차 초안 생성기(draft generator)로서 기능할 때 가장 현실적이고 이상적인 도구가 될 수 있다.

필자는 LLM 기술이야말로 동시대의 생동하는 소셜미디어 언어와 교과서 중심의 교실 언어 사이의 오랜 간극을 좁힐 수 있는 획기적인 도구라고 생각한다. 그러나 그 가능성을 현실화하는 것은 기술 자체보다 이를 비판적으로 수용하고 교육적으로 가공하는 교사의 역량에 달려 있다고 본다. ‘打工人’의 자조적인 한숨이나 ‘躺平’의 무력감, ‘破防’의 감정이 녹아 있는 소셜미디어의 언어는 최근 중국 청년들의 삶을 비추는 거울이다. 이를 교실 안으로 들여오는 작업은 단순한 어휘 학습을 넘어, 학생들이 동시대 중국인들의 생생한 숨결과 문화를 만나는 통합적인 경험이자 체험이 될 것이다. 물론 시시각각 변하는 소셜미디어의 흐름을 좇아 매년 새로운 텍스트를 교육용으로 가공하는 일은 교사들에게 있어 수고롭고 번거로운 과제일 수 있다. 그러나 맥락이 살아있는 최신 주제와 어휘를 선별하여 학생들에게 전달하는 안목은 오직 교사만이 발휘할 수 있는 고유한 전문 역량이다. 본고의 ChineseRealTalk과 같은 보조 도구가 자료 준비의 부담을 획기적으로 덜어주는 만큼, 교육자들은 약간의 수고로움을 감수하더라도 생생

한 온라인 텍스트를 교실 안으로 적극적으로 도입하는 능동적인 큐레이터의 역할을 기꺼이 수행해야 할 것이다.

<참고문헌>

- 박민준, 〈웨이보에 투영된 疫情 - COVID-19 관련 키워드의 SNS 빅데이터 분석〉, 《중국언어연구》 제90집, 2020.
- 박민준, 〈중국 소셜미디어 크롤링과 텍스트 분석: SNS 말뭉치 기반 연구 방법론의 모색〉, 《시민인문학》 제46호, 2024.
- 송지현, 〈현대 중국어의 신조어와 유행어 유형분석〉, 《중국어문학논집》 제21호, 2002.
- 주강훈, 〈중국어 언어교육의 관점에서 바라본 중국 MZ세대의 어휘사용 현황 소고〉, 《중국문학연구》 제91집, 2023.
- 하두진·박민준, 〈텍스트마이닝을 활용한 京派와 海派 작품분석: RAG 기반 챗봇 분석 가능성과 한계점〉, 《시민인문학》 제49호, 2025.
- 李泉, 《對外漢語教學學科理論研究》, 北京: 商務印書館, 2006.
- 梁雲雲, 〈來華留學生漢語教學中的網絡流行語調查及教學策略探析〉, 《高教論壇》 第8期, 2021.
- 陸丹妮, 〈對外漢語教學中的網絡新詞教學策略探析〉, 《語文學刊》 第2期, 2015.
- 朴敏浚, 〈面向初學者的中文信息處理平台構建及應用〉, 《數字化漢語教學》, 北京: 清華大學中文系, 2016.
- 朴敏浚, 〈基於語料庫的中文在線仇恨言論的統計研究〉, 《中國語文學論集》 155號, 2025.
- 顏圭源, 《lxy: 中文互聯網網絡流行語數據倉庫》, GitHub repository, 2023.
<https://github.com/yanguiyuan/lxy>

鄭金芝, 〈對外漢語網絡新詞新語教學策略研究〉, 《揚州教育學院學報》 40卷 1號, 2022.

Drkameleon, *Complete HSK Vocabulary*, GitHub repository, 2023. <https://github.com/drkameleon/complete-hsk-vocabulary>.

Park Minjun, Lee Seulki, “From Plain to Hierarchical: Knowledge-Augmented Prompting for Chinese Factivity Inference”, *Proceedings of CCL25-Eval Task 4*, 2025.

Sarhi Parth, et al., “Raptor: Recursive abstractive processing for tree-organized retrieval”, *The Twelfth International Conference on Learning Representations*, 2024.

Zhao Wayne Xin, et al., “A survey of large language models”, *arXiv preprint arXiv:2303.18223* 1, no. 2, 2023.

<Abstract>

This study explores a methodology for converting Chinese social media texts into educational materials using Large Language Models (LLMs) and proposes an interactive platform called *ChineseRealTalk*. In Chinese language education, there has been a persistent gap between the language presented in textbooks and the authentic everyday language used by native speakers on social media platforms. Despite the pedagogical value of social media texts that reflect the linguistic sensibilities of contemporary learners, incorporating such materials into the classroom remains challenging, as manually adapting them to match learners' proficiency levels and preparing customized learning texts require considerable time and effort.

To address this issue, this study developed *ChineseRealTalk* by leveraging the contextual understanding and text generation capabilities of LLMs. The platform offers three core functions: *Level-based Text Converter*, which rewrites original texts according to HSK levels; *Neologism and Difficult Expression Annotator*, which provides pinyin and example sentences for unfamiliar expressions; and *Comprehension Check Question Generator*, which automatically produces multiple-choice quizzes and discussion topics.

While the platform significantly reduces teachers' preparation time, the study also identifies several limitations, including LLM hallucinations, the loss of cultural context, and delays in reflecting the latest linguistic trends due to knowledge cutoff. Ultimately, LLMs function most effectively as draft generators for instructional materials. However, their effective use depends on teachers' active curatorial competence, including the critical evaluation of generated outputs and the careful design of prompts.

Key Words : Large Language Model (LLM), Social Media Text, Internet Slang, Chinese Language Education, Customized Text, ChineseRealTalk

